جامعة مؤتة

كلية الدراسات العليا

المقارنة بين الطريقة الحدية والطرق الارتباطية في حساب تمييز الفقرة في ضوء تباين طول الاختبار وحجم العينة

إعداد الطالب

اخلاص جميل خلف الصرايرة

إشراف

الدكتور راجي عوض الصرايرة

رسالة مقدمة إلى كلية الدراسات العليا

استكمالاً لمتطلبات الحصول على درجة الماجستير في

القياس والتقويم/ قسم علم النفس

جامعة مؤتة 2017

الآراء الواردة في الرسالة الجامعية لا تعبر بالضرورة عن وجهة نظر جامعة مؤتة

بسم الله الرحمن الرحيم

MUTAH UNIVERSITY
College of Graduate Studies

جامعة مؤتة
كلية الدراسات العليا

نموذج رقم (١٤)

قرار إجازة رسالة جامعية

تقرر إجازة الرسالة المقدمة من الطالبة اخلاص جميل الصرايره الموسومة بـ:

المقارنة بين الطريقة الحدية والطرق الارتباطية في حساب تمييز الفقرة في ضوء تباين طول الاختبار وحجم العينة

استكمالاً لمتطلبات الحصول على درجة الماجستير في **القياس والتقويم.**

القسم: علم النفس.

	التوقيع	التاريخ	
د. راجي عوض الصرايرة		٢٠١٧/٦/٢٠	مشرفاً ورئيسا
أ.د. فؤاد طه الطلافحة		٢٠١٧/٦/٢٠	عضواً
د. احمد عبدالله الطراونة		٢٠١٧/٦/٢٠	عضواً
د. حابس سعد الزبون		٢٠١٧/٦/٢٠	عضواً

عميد كلية الدراسات العليا

أ.د.محمد عبدالرحيم المحاسنه

MUTAH-KARAK-JORDAN
Postal Code: 61710
TEL :03/2372380-99
Ext. 5328-5330
FAX:03/ 2375694
sedgs@mutah.edu.jo dgs@mutah.edu.jo e-mail:
http://www.mutah.edu.jo/gradest/derasat.htm

مؤته – الكرك – الاردن
الرمز البريدي :٦١٧١٠
تلفون :٠٣/٢٣٧٢٣٨٠-٩٩
فرعي 5328-5330
فاكس 375694 ٠٣/٢
البريد الالكتروني
الصفحة الالكترونية

الإهداء

إلى زوجي............

إلى والد زوجي................

إلى والدي ووالدتي...........

إلى من تطلعوا لنجاحي بنظرات الأمل ... إلى إخوتي ....أخواتي

إلى من أسال أن يجعلهم الله قرة عين لي في الدنيا والآخرة وسترا من النار ... أبنائي وبناتي.

إلى كل من اسهم في اعداد هذا الجهد المتواضع

الشكر والتقدير

الحمد لله رب العالمين والصلاة والسلام على خير البشرية محمد عليه افضل الصلوات .

القائل: " من لا يشكر الناس لا يشكر الله "

لا يسعني الا الوقوف والتقدم بوافر الشكر والتقدير إلى استاذي الفاضل الدكتور راجي الصرايرة على ملاحظاته القيمة وعلى الجهود العلمية الكبيرة التي بذلها في مراحل اعداد الرسالة وكان لجهوده هذه الأثر البالغ في اخراج الرسالة بالشكل التي هي عليه الأن .

كما اسجل خالص شكري وتقديري لأساتذة علم النفس التربوي / القياس والتقويم وزملائي طلبة الماجستير في تخصص القياس والتقويم.

فهرس المحتويات

المحتوى	الصفحة
الإهداء	أ
الشكر والتقدير	ب
فهرس المحتويات	ج
قائمة الجداول	هـ
الملخص باللغة العربية	و
الملخص باللغة الإنجليزية	ز
الفصل الأول : خلفية الدراسة وأهميتها	1
1.1 مقدمة	1
2.1 مشكلة الدراسة	3
3.1 أهمية الدراسة	3
4.1 أهداف الدراسة	4
5.1 أسئلة الدراسة	4
6.1 حدود الدراسة	5
7.1 مفاهيم الدراسة وتعريفاتها الاجرائية	5
الفصل الثاني : الاطار النظري والدراسات السابقة	6
1.2 الاطار النظري	6
1.1.2 الخصائص السيكومتيرية للاختبار	6
2.1.2 تحليل الفقرات	10
2.2 الدراسات السابقة	18
3.2 التعقيب على الدراسات السابقة	22
الفصل الثالث : المنهجية والتصميم	23
1.3 منهجية الدراسة	23

المحتوى	الصفحة
2.3 مجتمع الدراسة	23
3.3 عينة الدراسة	23
4.3 أداة الدراسة	24
5.3 إجراءات الدراسة	26
الفصل الرابع: عرض النتائج ومناقشتها والتوصيات	27
1.4عرض نتائج الدراسة ومناقشتها	27
2.4 التوصيات	31
المراجع	32
الملاحق	36

قائمة الجداول

الرقم	عنوان الجداول	الصفحة
.1	معاملات تمييز وصعوبة فقرات الاختبار	25
.2	الوسط الحسابي لمعامل التمييز باختلاف طول الاختبار وحجم العينة	28
.3	التحليل الوصفي للبيانات	29
.4	نتائج تحليل التباين الثنائي للكشف عن دلالة الفروق بين متوسطات معامل التمييز	30

الملخص

المقارنة بين الطريقة الحدية والطرق الارتباطية في حساب تمييز الفقرة في ضوء تباين طول الاختبار وحجم العينة.

اخلاص جميل الصرايرة

جامعة مؤته، 2017

هدفت هذه الدراسة إلى المقارنة بين الطريقة الحدية والطرق الارتباطية في حساب تمييز الفقرة في ضوء تباين طول الاختبار وحجم العينة، وللإجابة عن أسئلة الدراسة تم استخدام بيانات اختبار محكي المرجع في مبحث الرياضيات للصف التاسع تكون من (35) فقرة من نوع الاختيار من متعدد بأربعة بدائل بعد التأكد من صدق الاختبار وثباته بالطرق الاحصائية المناسبة، حيث تم استخدام ثلاث احجام للعينة تم اخذها بالطريقة العشوائية وهي (150، 100، 50) مفحوص، وكذلك ثلاثة اطوال للاختبار اختيرت بطريقة عشوائية وهي (35، 25، 15) فقرة، وتم حساب معاملات التمييز تبعاً لطول الاختبار وحجم العينة بثلاث طرق حسابية وهي (الطريقة الحدية، معامل ارتباط بوينت بايسيريال، معامل ارتباط فاي)، حيث تم حساب التمييز بالطريقة الحدية وطريقة معامل الارتباط بوينت بايسيريال باستخدام برمجية الاكسل وفقا للصيغ الرياضية لكل طريقة، أما طريقة معامل فاي فتم حساب التمييز باستخدام برمجية (spss)، وللمقارنة بين الطرق المستخدمة تم حساب المتوسط الحسابي للتمييز لكل طريقة وفقاً لطول الاختبار وحجم العينة، كما تم استخدام تحليل التباين الثنائي لمعرفة أثر حجم العينة وطول الاختبار والتفاعل بينهما على التمييز، واشارت النتائج إلى عدم وجود اثر معنوي في حساب التمييز في ضوء تباين حجم العينة وطول الاختبار والتفاعل بينهم وفقا لقوانين النظرية الكلاسيكية في القياس لكل طريقة على حدا.

Abstract

Comparison of Marginal Method and Relational Methods in Calculating the Highlight the Item in the light of the Varying Length of the Test and Sample Size.

Ekhlass Jamil Alsaraireh

Mu'tah University, 2017

The objective of this study was to compare the marginal way and relational methods in calculating the highlight the item in the light of the length of the test sample size variation, and to answer the study questions spoken test data is used, the reference in the subject of mathematics to ninth grade (35) item of multiple choice type with four substitutes after confirming the veracity of the test and the realbilits of appropriate statistical methods have been using three sample sizes were considered indiscriminate manner (150, 100, 50) overhauled, As well as three lengths to test selected at random (35, 25, 15), and the discrimination index coefficients depending on the length of the test and sample size calculation in three ways (the marginal way, baisirial point correlation coefficient, coefficient of correlation phi) where marginal way discrimination is calculated and how the correlation coefficient point baisirial using Excel according to mathematical formulas for each method, Either way it was a factor of discrimination account using software (spss), and compare the methods used are average for all manner of discrimination index according to the length of the test and sample size, and get the results analysis of variance was used to determine the impact of sample size and length of the test and their interaction on the discrimination index, The results indicated a lack of moral effect on account of discrimination index in the light of the varying sample size and length of the test and the interaction between them according to the laws of classical theory in the measurement of each method.

الفصل الأول

خلفية الدراسة وأهميتها

1.1 مقدمة:

تشكل قضية تطوير الاختبارات التحصيلية، مكانة متميزة في مجال علم النفس بعامة، والقياس النفسي بخاصة ويرجع ذلك إلى كونها من أهم الوسائل المستخدمة في مجال التقويم، الأمر الذي يتطلب إعداد وبناء اختبارات نفسية موثوقًا في كفاءتها تستخدم في مجالات مختلفة، وإذا ما انتقلنا إلى مجال قياس الذكاء سنجد أن هذا المجال يمثل حجر الزاوية في مجال القياس النفسي، ونتيجة لتطور القياس النفسي والذي انعكس بدوره على تطور الاختبارات والمقاييس كان الهدف الذي يسعى إليه علماء القياس النفسي هو تحقيق موضوعية القياس ؛ لذا شهد هذ ا الميدان تطورات متزايدة تتعلق بأساليب تصميم وبناء وتحليل فقرا ت الاختبار لتحقيق هذا الهدف (ابو حطب، 1992).

ويميز القياس النفسي والتربوي بين مدخلين رئيسين في تصميم وبناء الاختبارات والمقاييس وتحليل البيانات المستمدة منها، هما :

المدخل التقليدي المتمثل في النظرية الكلاسيكية للاختبارات (Classical Test Theory (CTT) وما تنطوي عليه من مفاهيم ومبادئ بعضها يتعلق بخصائص فقرات الاختبار الصعوبة، والتمييز، والبعض الآخر يتعلق بخصائص الاختبار ككل (الصدق، والثبات، والمعايير) (علام، 2001).

المدخل المعاصر والمتمثل في نظرية السمات الكامنة ((Latent Trait Theory (LTT) . أو نظرية الاستجابة للفقرة كما يسميها بعض علماء القياس أمثال لورد (Lord)، (Hambleton) هامبلتون لاهتمامها بالربط بين استجابة الفرد لفقرة اختبارية وبين خصائص هذه الفقرة (علام، 2005).

ولقد كانت النظرية الكلاسيكية للاختبار لسنوات عديدة مضت، هي الأساس النظري والعملي للقياس في العلوم السلوكية . والتي استندت على نموذج الدرجة الحقيقية للفرد وهي أن لكل فرد قدر ما من السلوك غير الملاحظ (الدرجة الحقيقية True Score) والذي لا يمكن ان يقاس بصورة مباشرة، وعلية فان عدد الاجابات

الصحيحة الملاحظة هي التي تقيس نفس قدرة الفرد على اساس ان هذا القياس يتضمن قدراً من الخطأ اثناء تقدير الدرجة الحقيقية (عبدالحافظ ، 1999).

فدرجة الفرد الملاحظة في الاختبار(X) تساوي مجموع الدرجة الحقيقية للفرد في السمة المقاسة (T) مضافاً إليه الدرجة الخطأ في الاداء على الاختبار (E) التي قد تكون موجبة أو سالبة أو كلاهما.

وقد اشار كروكر والجينا (Crocker& Algina,1986) ان نظرية القياس الكلاسيكية تعتمد على ثلاث افتراضات أساسية هي:

- الافتراض الأول: التوزيع الاعتدالي للدرجات على متصل القدرة التي يقيسها الاختبار، فعدد الأفراد الواقع في مستوى معين من القدرة، يناظر العدد المتوقع من دالة الكثافة الاحتمالية.
- الافتراض الثاني: أن الدرجة الملاحظة (X) هي حاصل جمع الدرجة الحقيقية زائدً خطأ القياس(E)، حيث أن (E) تأخذ قيماً سالبة و موجبة $X=T+(\pm E)$، وأن تباين الدرجات الخام يساوي تباين الدرجات الحقيقية مضافًا إليه تباين درجات الخطأ . كما أن الارتباط بين مجموعتين من الدرجات الخام، المستمدتين من اختبارين متوازيين، أو ثبات الاختبارات في مجتمع معين، يساوي تباين الدرجات الحقيقية إلى تباين الدرجات الخام، وأن الزيادة أو النقص في طول الاختبار، يؤثر في ثبات درجاته .
- الافتراض الثالث: أنه لا يمكن معرفة أو قياس الدرجة الحقيقية، بل يمكن الاستدلال عليها أو تقديرها من خلال الدرجة الملاحظة، وذلك بحساب متوسط الدرجة الملاحظة المستقلة، الناتجة من تطبيق الاختبار لعدد كبير من المرات، وتزداد دقة الدرجة الحقيقية بزيادة مرات التطبيق .

ومن هنا جاءت الحاجة لهذه الدراسة للمقارنة بين الطريقة الحدية والطرق الارتباطية لحساب تمييز الفقرة في ضوء تباين حجم العينة وطول الاختبار نظرا لأهمية الاختبارات التحصيلية التي يتم بناء عليها اتخاذ القرارات وتوجيه عملية التدريس للحصول على نتائج دقيقة مبنية على اسس علمية واحصائية دقيقة في حساب معامل تمييز الفقرة بالطرق المختلفة والتوصل إلى

نتائج الدراسة من خلال المقارنة بين هذه الطرق الارتباطية والطريقة الحدية وصولاً إلى نتائج الدراسة.

2.1 مشكلة الدراسة:

تواجه الاختبارات التحصيلية في مجال القياس النفسي والتربوي العديد من المشكلات التي تقلل من مصداقيتها، ولعل أبرز هذه المشكلات هي:

أولاً: تأثر الدرجة الكلية للمفحوص بخصائص فقرات الاختبار مثل الصعوبة والتمييز والثبات والخطأ المعياري فقيمة صعوبة فقرة ما، تعتمد على مستوى المجموعة التي تقدمت للاختبار لذلك سترتفع قيمتها إذا كانت المجموعة ذات قدرة عالية وستنخفض قيمتها إذا كانت المجموعة أقل قدرة، كما سيختلف معامل التمييز في حال تجانس أو عدم تجانس قدرات المفحوصين، وبمعنى آخر فإن تقدير معالم الفقرة ليس مستقلا عن العينة التي يطبق عليها.

ثانياً: قصور بعض الاختبارات في التعبير عن القدرة الحقيقية للمفحوص.

ثالثاً: القصور في وجود محك ثابت لمقارنة الدرجة التي حصل عليها الطالب في الاختبار بهذا المحك.

رابعاً: عدم وجود تفسير واضح وثابت للدرجات على الاختبار واعتماد ذلك على مجتمع الأفراد الذي أعد له الاختبار، وعلى ظروف تطبيق الاختبار.

لذلك تكمن مشكلة هذه الدراسة في البحث في أثر حجم العينة وطول الاختبار على تقدير معامل تمييز الفقرة في تقديرها من خلال اختبار تحصيلي في الرياضيات في ضوء مؤشرات التمييز في النظرية الكلاسيكية للاختبار.

3.1 أهمية الدراسة:

تنبع أهمية الدراسة في جانبين أولهما الجانب النظري والذي يتمثل في تقديم الادب النظري عن الطرق الحسابية المختلفة لحساب معامل التمييز للفقرة بالطريقة الحدية والطرق الارتباطية، وثانيهما هو الجانب العملي وتمثل ذلك في اعطاء نتائج تتعلق في المقارنة بين الطريقة الحدية والطرق الارتباطية في حساب تمييز الفقرة في

ضوء تباين طول الاختبار وحجم العينة لمعرفة الطرق المستخدمة لحساب تمييز الفقرة كما سيظهر معنا في نتائج الدراسة، كون ان هذه الطرق تستخدم بشكل كبير في البحوث والدراسات وجاءت هذه الدراسة لتقيم فعالية هذه الطرق في ضوء تباين حجم العينة وطول الاختبار.

4.1 أهداف الدراسة:

هدفت هذه الدراسة إلى المقارنة بين الطريقة الحدية والطرق الارتباطية في حساب تمييز الفقرة في ضوء تباين طول الاختبار وحجم العينة. وكذلك التعرف إلى قيم مؤشرات تمييز الفقرة باستخدام طرق الحساب المختلفة باختلاف طول الاختبار وحجم العينة وأثر التفاعل بينهما على حساب تمييز الفقرة .

و تسعى الدراسة إلى

1- التعرف إلى قيم مؤشرات تمييز الفقرة حسب طرق حساب مؤشر التمييز تبعاً لطول الاختبار.
2- التعرف إلى قيم مؤشرات تمييز الفقرة حسب طرق حساب مؤشر التمييز تبعاً لحجم العينة.
3- التعرف إلى قيم مؤشرات تمييز الفقرة حسب طرق حساب مؤشر التمييز تبعاً للتفاعل بين طول الاختبار وحجم العينة.

5.1 أسئلة الدراسة:

تتمحور أسئلة الدراسة بالآتي:

السؤال الأول: هل تختلف قيم مؤشرات تمييز الفقرة حسب طرق حساب مؤشر التمييز تبعاً لطول الاختبار؟

السؤال الثاني: هل تختلف قيم مؤشرات تمييز الفقرة حسب طرق حساب مؤشر التمييز تبعاً لحجم العينة؟

السؤال الثالث: هل تختلف قيم مؤشرات تمييز الفقرة حسب طرق حساب مؤشر التمييز تبعاً للتفاعل بين طول الاختبار وحجم العينة؟

6.1 حدود الدراسة:

1-الحدود الزمانية: تم جمع البيانات في الفصل الثاني من العام الدراسي (2017/2016).

2-الحدود المكانية: أجريت هذه الدراسة في مدارس مديرية التربية والتعليم قصبة الكرك.

3-الحدود الموضوعية: ينحصر موضوع الدراسة الحالية في مقارنة الطريقة الحدية بالطرق الارتباطية لحساب تمييز الفقرة في ضوء تباين طول الاختبار وحجم العينة.

4-الحدود البشرية: تستهدف الدراسة طلاب الصف التاسع الاساسي لمبحث الرياضيات في مدارس قصبة الكرك.

7.1 مفاهيم الدراسة:

1. الطريقة الحدية: تشير إلى طريقة حساب معامل التمييز للفقرة ويكون في هذه الدراسة عبارة عن نسبة الذين أجابوا إجابة صحيحة من الفئة العليا مطروح منها نسبة الذين أجابوا إجابة صحيحة من الفئة الدنيا فيكون الناتج قيمة معامل التمييز وتعد من اسهل طرق الحساب لمعامل التمييز.
2. الطرق الارتباطية: تتمثل في استخدام معاملات الارتباط لحساب معامل التمييز وفي هذه الدراسة سيتم استخدام طريقتين هما (معامل ارتباط فاي، ومعامل ارتباط بوينت بايسيريال).
3. تمييز الفقرة: يتمثل في تحديد مدى فاعلية فقرة ما في التمييز بين الطالب ذي القدرة العالية والطالب الضعيف بالقدر نفسه الذي يفرق الاختبار بينهما في الدرجة النهائية بصورة عامة.
4. طول الاختبار: يشير إلى عدد فقرات الاختبار التي سيتقدم لها المفحوص.
5. حجم العينة: تمثل عدد المفحوصين أو المتقدمين للاختبار.

الفصل الثاني

الإطار النظري والدراسات السابقة

1.2 الإطار النظري:

إن مستوى اداء المفحوص في الاختبار يتوقف على خصائص المفحوص والاختبار (Hambleton,1983). لذلك فإن معامل صعوبة الفقرة ومعامل تمييزها يؤثر على اداء الافراد. وصعوبة الفقرة احصائيا هي نسبة الافراد الذين أجابوا إجابة صحيحة على الفقرة. أما صعوبة الفقرة المعرفية فالمقصود بها طريقة المعالجة العقلية التي تتطلبها المهمة وتعرف إجرائياً بالمستوى المعرفي الذي تقيسه تلك الفقرة (Lane and Bull,1987). أما بالنسبة للقدرة التمييزية للفقرة فيقصد بها قدرة الفقرة على التمييز بين الافراد ذوي الاداء المرتفع والافراد ذوي الاداء المنخفض، واحصائياً هو الفرق بين عدد الطلاب الذين أجابوا إجابة صحيحة من الفئة العليا وعدد الطلاب الذين أجابوا إجابة صحيحة من الفئة الدنيا على عدد افراد احدى الفئتين، ولا تقتصر الخصائص المؤثرة على الاداء على صعوبة وتمييز الفقرات بل تشمل ايضاً ثبات الاختبار وصدق الاختبار وترتيب فقرات الاختبار (Plake et.al,1981).

1.1.2 الخصائص السيكومترية للاختبار.

نجد ان مصمموا الاختبار يهتمون بأن يمتاز الاختبار بدرجة عالية من الصدق والثبات كي يصار إلى اعتماده بثقة وقوة. ولما كان الصدق والثبات من الركائز الاساسية في الاختبار كان لا بد من إيلائه الاهمية القصوى كي تأتي النتائج على مستوى عال من الثبات والتقدير(خوري، 2008).

أولاً: الصدق، وهذه خاصية للاختبار نفسه، وتستخدم علامات المفحوصين على ذلك الاختبار في الاستدلال عن هذه الخاصية. ويعرف الصدق بانه الدقة التي يقيس بها الاختبار ما يجب قياسه، أو ما مدى تأدية الاختبار للوظيفة التي استخدم من أجلها، أو مدى قيامه بالوظيفة المفترض قيامه بها(النجار، 2010).

فقد ذكر الحجاج(2013) بان الصدق عبارة عن الدرجة التي يتمتع بها الاختبار

من تحقيق اهداف محددة وقد اكد على اهداف عامة كالدليل على صدق الاختبار وهي:

1-التنبؤ في أمور ذات علاقة بالمفحوص

2-وصف المفحوص من جانب أو مجموعة من الجوانب

يرتبط صدق الاختبار بصحته وصلاحيته للاستخدام، فالاختبار الصادق هو الاختبار الذي يصلح للاستخدام في ضوء الاهداف التي وضع من اجلها.

للصدق اهمية كبيرة لتحديد قيمة الاختبار، فعنوان الاختبار قد يكون عبارة صغيرة لا تدل على ما يقيسه بشكل عام، فالصدق الذي يقرر صلاحية أي فرض علمي أو في تحديد مفهوم معين، حيث لخصت اهمية الصدق بما يلي (النجار، 2010):

1- تحديد لحقيقة اداء الفرد في الوقت الحاضر في نطاق شامل لمواقف سلوكية.
2- التنبؤ بالأداء المستقبلي للفرد أو تقدير مكانه في احد المتغيرات ذات الاهمية.
3- الاستدلال على درجة تملك الفرد لسمه أو خاصية معينة من حيث هي تكوين فرضي.

ويمكن اجمال معاني صدق الاختبار على النحو التالي(عبدالهادي، 2002):

أ- الصدق مفهوم نسبي: أي أن الاختبار يمكن ان يكون صادقاً لقياس سلوك طلبة من صف ما، وليس صادقاً في قياس سلوك طلبة من نفس الصف في بيئة اخرى، فالصدق يعتمد على طبيعة الجماعة التي طبق عليها الاختبار، فالاختبار الذي يكون صادقاً في بيئة مدنية، قد لا يكون صادقا بالدرجة نفسها، اذا استخدمناه في بيئة ريفية، والاختبار الذي قد يكون صادقاً في قياس سلوك الطلبة الذكور، قد لا يكون صادقاً بالدرجة نفسها في قياس سلوك الإناث.

ب- الصدق النوعي: فالاختبار الصادق الذي يقيس معارف الطلبة في مادة التربية البيئية، لا يكون صادقاً إذا استخدمناه في قياس سلوك اخر كالاتجاهات البيئية على سبيل المثال، والمهارات الحركية الرياضية، ولا

يكون صادقا ألا إذا استخدمناه في قياس السلوك الذي اعد من اجل قياسه.

ج- مفهوم الصدق يتضمن الثبات: الاختبار الصادق يقيس ما اعد لقياسه بدقة، فاذا استخدم معلم اللغة العربية اختباراً صادقاً لقياس سرعة القراءة لدى طلبة الصف الأول الاساسي، فان هذا الاختبار لن يكون صادقاً اذا اعطى نتائج مختلفة في حال استخدامه مرة ثانية في قياس السلوك نفسه، فالاختبار الصادق هو اختبار ثابت ايضاً في نتائجه.

د- الصدق ليس درجة مطلقة أو صدق تام: فالاختبار الصادق ليس صادقا بدرجة كاملة، إذ ترتفع درجة صدق الاختبار لقياس الذكاء (90%) ويسمى الاختبار صادقاً، وقد تصل صدق الاختبار لقياس الذاكرة إلى(70%) ويسمى الاختبار صادق.

ومن انواع الصدق المرتبط بمحك:

أولاً: الصدق التنبؤي (predictive validity). في هذا النوع يتم جمع البيانات على المحك بعد فترة طويلة نسبياً من جمع البيانات على الاختبار أو اداة القياس التي نبحث عن صدقها، ونحتفظ بنتائجهم على الاختبار إلى ان يخضعوا للمحك في فترة زمنية لاحقة، ونقوم بحساب معامل الار تباط بين نتائجهم على الاختبار والمحك.

ثانياً: الصدق التلازمي (concurrent validity). ويتم في هذا النوع من الصدق جمع المعلومات على الاختبار الذي نبحث عن صدقه، وعلى المحك في نفس الفترة الزمنية، أو بفاصل زمني قصير جداً، وبما ان جمع المعلومات متزامن على المتغيرين (المتنبئ والمحك) فان اي منهما يمكن ان يحل مكان الاخر دون تمييز، وخاصة عندما يصعب ترجيح اي منهما كمحك.

ثانياً: الثبات (reliability).

يقصد بثبات الاختبار، أن علامة الفرد لا تتغير جوهرياً بتكرار أجراء الاختبار ويعبر عنه احصائياً بانه معامل الارتباط بين علامات الافراد وبين مرات أجراء الاختبار المختلفة، أن الاختبار يصبح موثوق به ويعتمد عليه، فاذا تم عمل أي اختبار تحليلي بصورة دقيقة فان العلامات الصحيحة والتي ليس فيها

خطأ، تعزى إلى مجموعة الاسئلة الخاصة التي استعملت في الاختبار، وهذا يتطلب صياغة اسئلة الاختبار بطريقة معينة، وإلى ظروف خاصة تمنع التخمين والقلق والفوضى والعوامل الاخرى التي يمكن ان تؤثر على انجازات المفحوصين وينتج عن ذلك اختبار ذو ثبات عالي، واذا كان ثبات الاختبار قليل فانه لا ينجح في قياس الشيء بدقة(الظاهر واخرون، 2002).

يتعلق ثبات الاختبارات الصفية بمدى دقة هذه الدرجات كدليل أو مؤشر للتحصيل الفعلي للطلاب. إن درجات الاختبار الصفية ينبغي ان تعكس الفروق بين الطلاب في تحقيقهم لأهداف المواد الدراسية التي تقيسها هذه الاختبارات (علام، 2007). فقد اشار (عودة، 2010) إلى انواع معاملات الثبات التالية:

انواع معاملات الثبات

أولا: معامل ثبات الاستقرار (stability).

قد ذكر عودة (2010) أن هناك نوعان من الصدق لكي تكون الاختبارات الصفية مفيدة في موقف معين، ينبغي ان تكون درجاتها التي يحصل عليها الطلاب متسقة وموائمة لهذا الموقف. فالاختبار يمكن ان ينتج عنه درجات متسقة، وبذلك تتميز الدرجات بالثبات ولكن الاختبار نفسه ربما يكون عديم الفائدة في مواقف اخرى، حيث انه ربما لا يفي بمتطلبات هذا الموقف. فالاختبار الذي يقيس قياساً متسقاً لما هو موائم يعد اختباراً صادقاً، فالثبات يعد شرطا ضرورياً لصدق درجات الاختبارات، ولكنه ليس كافياً. اذ يمكن ملاحظة سلوك ما باتساق كبير، ومع ذلك تكون هذه الملاحظة غير صادقة في اتخاذ القرار أو التوصل إلى استدلال معين لذلك فانه على الرغم من اهمية الثبات الا ان الصدق يعد اكثر اهميه (علام، 2007).

ثانياً: معامل ثبات التكافؤ(equivalence).

تقوم فكرة هذا المعامل على حساب الارتباط بين علامات عينة من الطلبة على الاختبار المرغوب حساب ثباته وعلاماتهم على اختبار اخر مكافئ للاختبار الأول ويقيس نفس السمة، وهنا يقصد بالتكافؤ ان الاختبارين لهما نفس الخصائص من حيث الصعوبة والقدرة التمييزية، وتشابه المحتوى الذي يقيسه الاختبار، وهناك شروط للتكافؤ

على مستوى الاختبار، الشرط الأول ان يكون نفس التوزيع للعلامات على الاختبارين، والشرط الثاني ان يكون لهما نفس الوسط الحسابي، ونفس التباين، وليس المقصود هنا التطابق في القيم، ولكن ان يكون الفرق غير دال احصائياً. على ان يطبق الاختبار على نفس المجموعة بنفس الوقت أو بفاصل زمني بسيط.

ثالثاً: معامل استقرار التكافؤ (stability-equivalence).

تقوم فكرة هذا المعامل على حساب الارتباط بين علامات مجموعة الاختبار على صورتين متكافئتين للاختبار مع وجود فاصل زمني طويل نسبياً بين تطبيق الصورة الأولى للاختبار والصورة الثانية، وذلك لتقليل اثر التذكر أو اثر التحمل، وهذا يعني ان هذه الطريقة مركبة من الطريقتين السابقتين.

رابعاً: معاملات ثبات الاتساق الداخلي.

جميع الطرق السابقة تناسب الاختبارات المقننة اكثر من الاختبارات التي يجريها المعلم، لأنها تحتاج إلى تطبيق الاختبار أو صورة مكافئة له اكثر من مرة، الا ان معاملات ثبات الاتساق الداخلي لا تتطلب تطبيق الاختبار الا مرة واحدة.

2.1.2 تحليل الفقرات.

ان تحليل فقرات الاختبار والاختبار بصورته الكلية ذات اهمية كبيرة وضرورية لبناء الاختبار، وذلك لان صحة الاختبار من صحة فقراته، فاذا ادت عملية تحليل فقرات الاختبار إلى زيادة صحتها وثباتها يؤدي ذلك إلى زيادة في صحة الاختبار بصورته الكلية(جرادات واخرون، 2002).

حيث ان التحليل الاحصائي يقدم تفسيرا للعلامات التي حصلنا عليها بعد تطبيق الاختبار، بحيث يتم توضيح درجات الطلبة بصورة كلية أو نسبية، ويتمثل ذلك بتحليل الدرجة على كل فقرة، وهذا يزودنا بمؤشرات التمييز التي تتمثل في التعرف على مستوى السهولة والصعوبة(عبدالهادي، 2002).

فقد ذكر كروكر ولجينا (Crocker& Algina,1986)، أن الهدف بشكل عام إلى الوصول لاختبار باقل عدد من الفقرات ويؤدي إلى درجات لها درجة مناسبة من الثبات والصدق للاستخدامات الخاصة بالاختبار. ويرافق هذا تجريب ميداني لعدد كبير من

الفقرات، واختبار مجموعة فرعية منها تساهم بأكبر قدر من ثبات الاختبار أو صدقه. وفي عملية بناء اختبار جديد (أو اختصار عدد فقرات) فان المجموعة الاخيرة من الفقرات تحدد عادة خلال عملية تسمى تحليل الفقرات. وتحليل الفقرات مصطلح واسع الاستخدام في تحديد الحسابات واختبار الكثير من الخصائص الاحصائية لاستجابات المفحوصين على فقرة اختبارية. ويتم التحقق من معالم الفقرات والتي تقع ضمن ثلاث فئات عامة هي:

1. معاملات تصف توزيع الاستجابات لفقرة مفردة(مثل المتوسط والتباين للاستجابات على فقرة)
2. معاملات تصف درجة العلاقة بين الاستجابات على الفقرة واي محك مناسب.
3. معاملات عبارة عن دوال لكل من تباين الفقرة وعلاقته بالمحك.

أذا افتراضنا عشوائية التخمين فان قيمة أ الملاحظة مثلاً يتوقع أنها مجموع نسبة المفحوصين الذين يعرفون الإجابة الصحيحة (قيمة أ الحقيقة) و $\frac{1}{m}$ من نسبة الذين لا يعرفون الإجابة الصحيحة ، حيث ترمز(m) إلى عدد الاختبارات أو بدائل الإجابة. ولتعظيم تباين الدرجة الحقيقية الكلي إلى عدد(m) فمن الضروري تعظيم تباينات الدرجة الحقيقة للفقرة .ويمكن تعظيم التباين عندما يجيب نصف المفحوصين على الفقرة اعتماداً على معرفتهم بينما النصف الاخر لا يستطيع ذلك، وعبر هذا النصف الذي لا يعرف الإجابة الصحيحة فان لكل مفحوص احتماليه = $\frac{1}{m}$ ليجيب على الفقرة إجابة صحيحة بالتخمين العشوائي من خلال البدائل التي عددها(m). لذا فان نسبة المفحوصين المتوقع ان يجيبوا إجابة صحيحة بالتخمين =$\frac{.50}{m}$.

وبالتالي فان الصعوبة الملاحظة للفقرة بأقصى تباين درجة حقيقية يتوقع ان يساوي :

$$P_0 = .50 + \frac{.50}{m}$$

وأثبت لورد (Lord ,1952) في دراسة محاكاة ان الثبات يتحسن عند اختيار فقرات لها قيم اكبر لمعامل الصعوبة(p) من تلك المحسوبة بالتعديل الناجم عن التخمين، وقيم(p) من جهة التطبيق العملي فان قيم لورد تعتمد على بيانات استجابة محاكاة (افتراضية وليست واقعية)، مع ذلك فإنها تبقى اكثر معقولية من تلك المحسوبة

بنموذج التخمين العشوائي ، لان العديد من المفحوصين لديهم معرفة جزئية تمكنهم من حذف واحد أو اكثر من البدائل الخاطئة قبل اجراء عملية التخمين، مما يجعل احتمالية الإجابة الصحيحة اكبر من النسبة $\frac{1}{m}$.

تمييز الفقرات

تقدم العديد من الاختبارات معلومات عن الفروق الفردية سواء على المقياس (الاختبار) أو بناء على محك خارجي يفترض ان تتنبأ به درجات الاختبار. وفي كلا الحالتين فان المعلم الذي نهتم به في اختبار الفقرات يجب ان يكون مؤشرا لفعالية تمييز الفقرة بين المفحوصين ذوي الاداء العالي وذوي الاداء المنخفض على المحك نفسه . ولا يوجد في بعض الحالات مقياس انسب للبناء من الدرجة الكلية على الاختبار نفسه (الاختبار التحصيلي الصفي هو مثال اساسي لهذا الموقف). وفي هذه الحالة تستخدم الدرجة الكلية كتعريف اجرائي للموقف النسبي للمفحوص على البناء قيد الاهتمام . وفي هذا المحك الداخلي يكون الهدف تحديد الفقرات التي تكون احتمالية الإجابة الصحيحة عليها عالية من قبل فئة التحصيل العالي. ومنخفضة في الوقت نفسه من قبل فئة التحصيل المتدني وفي الاختبارات التحصيلية على سبيل المثال نقول ان الفقرة تميز أو تفاضل بين المفحوصين الذين يعرفون المادة والذين لا يعرفونها من جهة اخرى فإننا نشك في الفقرة التي يتساوى فيها نجاح الفئتين . فمثل هذه الفقرات يبدو انها لا تقيس البناء نفسه الذي تقيسه بقية الفقرات . ومن غير المرغوب به ايضا ان يكون هنالك فقرات تفشل الفئة العليا في الإجابة عليها إجابة صحيحة وتنجح الفئة الدنيا في الإجابة عليها بصورة صحيحة، فمثل هذه الفقرات يكون تمييزها سالب.

وتستخدم خمسة معالم على انها مؤشرات لفعالية تمييز الفقرة، وسيتم وصفها في الجزء التالي. يعتمد احدها على التفاصيل بين مجموعات من المفحوصين تم تحديدهم من خلال درجات قطع عند نقاط محددة على توزيع درجات المحك، والاربعة المتبقية هي انواع مختلفة من معاملات الارتباط. ومن المهم تمييز ان هذه المعالم جميعها يمكن تطبيقها بالتساوي على المواقف بالأداء على محك خارجي. ومع ان التطبيق الاخير مثيرا للجدل، الا ان اختيار محك داخلي (الدرجة الكلية) لا يغير من الصيغ المحددة أو الطرائق الحسابية.

دليل التمييز:(Index of Discrimination)

هو معلم تمييز بسيط يطلق عليه اسم دليل التمييز يمكن تطبيقه فقط على الفقرات ثنائية التصحيح ويتطلب حسابه تحديد نقطة أو اثنتان على توزيع درجات المحك على انها درجات القطع، وتفصل المفحوصين إلى مجموعات درجاتهم اعلى أو ادنى من درجات القطع هذه. على سبيل المثال، ان كان اهتمام مطور الاختبار في اختيار الفقرات التي تمييز بين المفحوصين بناءً على محك داخلي (درجة الاختبار الكلية)، والمجموعات تتألف من اعلى (50%) وادنى(50%) من المفحوصين اعتمادا على درجاتهم الكلية، وكانت مجموعة المفحوصين كبيرة فليس من الضروري دراسة تقليدية استخدام الفئة العليا أو الدنيا جميعها (Kelley,1939) واجراها كيلي تحت شروط معينة، ووجد انه يمكن الحصول على معامل تمييز فقرة اكثر حساسية واستقراراً عند استخدام اعلى (27%) وادنى (27%) من المفحوصين وذلك عندما يكون حجم العينة كبيرا. وفي الحقيقة يمكن الحصول على النتائج نفسها من خلال اعلى (30%) أو (50%) من المفحوصين (Benchert & Mendoza ,1979;Englehart,1965) وحالما يتم تحديد الفئتين العليا والدنيا، فان دليل التمييز يحسب ب:

$$P_l - P_a = D$$

حيث ترمز P_a إلى نسبة الذين أجابوا على الفقرة إجابة صحيحة من الفئة العليا، و P_l هي النسبة نفسها للفئة الدنيا. وتتراوح قيم D بين (−1.00 إلى 1.00). وتشير القيم الموجبة إلى تمييز للفقرة يحابي الفقرة العليا من المفحوصين، وتشير القيم السالبة إلى تمييز عكسي للفقرة يحابي الفئة الدنيا.

المعاملات الارتباطية للتمييز:

درسنا ان معامل ارتباط بيرسون على انه مقياس لدرجة العلاقة الخطية بين متغيرين، فان كان مدى درجات فقرات الاختبار (1إلى 4) أو (1إلى5) أو اكثر من ذلك مثل فقرات استبانة اتجاهات ليكرت، فانه يمكن استخدام هذه الصيغة لتقدير درجة العلاقة بين درجات المحك.

معامل ارتباط بوينت بايسيريال:

عندما يكون مطور الاختبار مهتما بمعرفة مدى الارتباط بين الاداء على فقرة ثنائية التصحيح (درجاتها صفر، 1) ودرجة الاختبار الكلية (أو اي محك اخر تدريجه

متصل)، نستخدم هنا صيغة حسابية بسيطة لمعامل ارتباط بيرسون يسمى معامل ارتباط بوينت بايسيريال ويعرف على النحو التالي:

$$P_{pbis}= \frac{\mu+ -\mu_{(x)}}{\sigma_{\chi}}\sqrt{P_q}$$

حيث يرمز +μ إلى متوسط درجة المحك للذين أجابوا على الفقرة إجابة صحيحة، و $\mu_{(x)}$ إلى متوسط درجة المحك للمجموعة الكلية، و σ_{χ} إلى الانحراف المعياري، P إلى صعوبة الفقرة، و p-1=q.

وتكون قيمة معامل ارتباط بوينت بايسيريال نوعا ما اكبر من القيم الحقيقية، وذلك لان درجة الفقرة لها مساهمة في الدرجة الكلية لكل مفحوص. فان كان عدد الفقرات كبير نسبيا(25 أو اكثر) فان هذه الحقيقة نادرا ما تؤدي إلى مشكلة، وعليك تصحيح الصيغة بحل المشكلة للعدد القليل من الفقرات باستخدام الصيغة:

$$P1(x-i) \ = \frac{\boldsymbol{\sigma_i - \sigma_x - \sigma_{xi}}}{\boldsymbol{\sigma_i \sigma_x p_{xi} 2 - \sigma_x^2 + \sigma_i^2}}$$

حيث ترمز $P1(x-i)$ إلى معامل ارتباط درجة الفقرة بالدرجة الكلية بحذف الفقرة المعينة بمعامل الارتباط $\sigma_i \sigma_x$ *إلى الانحرافات المعيارية للدرجة الكلية ولدرجة الفقرة على التوالي*

معامل ارتباط بايسيريال:

عندما نرغب بان يكون توزيع المتغير الذي تقيسه الفقرة توزيعا اعتداليا، فمن الممكن اشتقاق صيغة للارتباط بين هذا المتغير ومحك توزيعه مثل درجة الاختبار. وقد اشتق هذا الاحصائي في البداية من قبل بيرسون (Pearson) واطلق عليه اسم معامل ارتباط بايسيريال، ويحسب باستخدام الصيغة:

$$P_{bis} = (Y/P)\ \frac{\mu_x - \mu_+}{\sigma_x}$$

حيث ترمز μ_+ إلى متوسط درجة المحك للذين أجابوا على الفقرة إجابة صحيحة، و μ_x هي متوسط درجة المحك للمفحوصين جميعهم و σ_x إلى الانحراف المعياري، P إلى نسبة الذين أجابوا إجابة صحيحة على الفقرة، و Y إلى الاحداثي Y للمنحنى الاعتدالي المعياري عند الدرجة الزائيه المرتبطة بقيمة P للفقرة.

هي: فان العلاقة بين معاملات ارتباط بايسيريال وبوينت بايسيريال

$$P_{pbis} = \sqrt{\frac{p_q}{y}}\ P_{bis}$$

ولان قيمة احداثيYعلى المنحنى الاعتدالي اقل من قيمة $\sqrt{P_q}$ فان قيمة معامل بايسيريال يكون على الاقل اكبر من ارتباط بوينت بايسيريال ب $\frac{1}{5}$ للمتغيرات نفسها (Lord&Novicck,1968) ويكون الفرق الكسري غير متحيز للفقرات متوسطة الصعوبة ، وبانخفاض قيمة P *إلى ما دون(0.25) أو ارتفاعها* اعلى من (0.75) فان الفرق بين معاملي الارتباط يزداد بحدة. وقد اثبت ماغنسون (Magnusson) بيانيا انه عند مدى الصعوبة المتطرف يكون معامل ارتباط بايسيريال اكبر اربع مرات من معامل ارتباط بوينت بايسيريال بين درجة الفقرة والدرجة الكلية. ويظهر هذا بسبب انخفاض اثر تباين الفقرة الذي يعمل على تقييد قيمة معامل ارتباط بوينت بايسيريال. لذا فمستخدمي الاختبارات الذين يقارنون بين تقارير تحليل الفقرات للدراسات المختلفة التي تستخدم صيغ معاملات ارتباط مختلفة عليهم تذكر ان معامل ارتباط بايسيريال اكبر وبانتظام من معامل ارتباط بوينت بايسيريال، وبالتالي فان هنالك فروقات واضحة في حساب قيم معالم تمييز الفقرة تعود إلى الصيغة المستخدمة في معامل الارتباط إلى الفروقات النوعية بين الفقرات.

معامل ارتباط فاي:

عندما نريد حساب معامل ارتباط فقرات ثنائية التصحيح ودرجات محك ثنائي(مثل ناجح وراسب في برنامج اعادة التأهيل، أو خصائص ديموغرافية مثل (الجنس) يمكننا استخدام معامل فاي. واستخدام اخر لهذا المعامل هو تحديد درجة الاستقرار في الاستجابات للفقرة نفسها ثنائية التصحيح للمفحوصين نفسهم في موقف اخر مختلف. ويكون استخدام معامل فاي مناسبا اكثر عندما تكون المتغيرات ثنائية حقيقية. فعند تشكيل مجموعات محكية باصطناع درجة قطع على توزيع المتصل، فان هذا الاحصائي لا يسمح باستخدام تام للمعلومات المناسبة. بكلمات اخرى عند تصنيف المفحوصين الذين اعلى من درجة القطع بالدرجة 1 وادنى منها بالدرجة صفر، فان المعلومات الكمية حول الفروق بين الدرجات في النجاح والفشل ستضيع. ومحدد اخر محتمل لمعامل فاي هو ان قيمته تساوي pعندما تكون قيمة أ لكلا المتغيرين متساوية

فقط (وذلك لأنه مثل معامل ارتباط بايسيريال النقطي يشتق مباشرة من معامل ارتباط بيرسون).

معامل ارتباط تتراشورك:

عندما يهتم مطور الاختبار بدرجة الارتباط بين متغيرين ثنائيين عندما يكون كلاهما ناجما عن تحويل متغير متصل إلى متغير متقطع ثنائي. و من المناسب في مثل هذه الحالات استخدام معامل ارتباط تتراشورك. وحساب هذا الاحصائي معقد جدا ونادرا ما يتم حسابه الا في حالات الاعتقاد القوي به من قبل مطور الاختبار اكثر من اي معامل اخر مثل معامل فاي الذي لا يكون مناسبا للهدف. وأحد هذه المواقف في حالة اجراء تحليل عاملي وتكوين مصفوفة متساوية، فان مثل هذه الارتباطات تكون اقل ملائمة من معاملات تتراشورك للتحليل العاملي. ولا تعرض صيغة معاملات تتراشورك بسهولة، ولكن يتوافر برامج محوسبة في العديد من الحقائب المحوسبة لحساب هذا المعامل .

مقارنة بين معاملات تمييز الفقرة:

تم عرض خمسة طرائق مختلفة في حساب تمييز الفقرة، ومن الواضح في بعض المواقف بسبب تصحيح المتغيرات، فان احدى التقنيات تكون اكثر ملائمة من غيرها. ومن المنطق طرح سؤال حول تشابه النتائج المحصلة من الطرائق المختلفة. وتناولت عدة دراسات تجريبية هذه القضية فقد استخدم انجل هارت(Englehart,1965) استجابات (210) مفحوصين تقدموا لاختبار كفاءة المدرسة العليا وباستخدام الدرجة الكلية كمحك، اجرى حساب معاملات تمييز متنوعة بما فيهاD، فاي، بايسيريال ، *بوينت بايسيريال ، وتتراشورك.* وعندما حسبت معاملات الارتباط بين هذه المعاملات وتراوحت الارتباطات بين ازواج معاملات التمييز بين (0.85) إلى (0.99) ومن (0.90) إلى (0.99) على الصيغة الموازية ودونت دراسات مشابهة من قبل Beuchert) &Mendiozaa , 1979 , Oosterhof ,1976 , Findley , 1956) في معظم الفجوات الاكبر للفقرات متطرفة الصعوبة.

وكخلاصة، يمكن تقديم التوصيات المتعلقة باختيار طريقة تمييز الفقرة للفقرات ثنائية التصحيح الاتية:

1. عندما تكون الفقرات متوسطة الصعوبة يكون الفرق بين احصائيات التمييز قليل، فان كان المهم سهولة الحساب ينصح باستخدام الطريقة الحدية (D) وان كان الاهتمام في اختبار الدلالة الاحصائية ينصح باستخدام احدى الطرائق الارتباطية.

2. وان كان الهدف اختيار الفقرات متطرفة الصعوبة يوصي باستخدام معامل ارتباط بايسيريال فيما لو كان من المنطقي افتراض التوزيع الاعتدالي للسمة المؤثرة في الاداء.

3. عندما يشك مطور الاختبار باختلاف قدرة العينات المستقبلية عن عينة تحليل الفقرات، يوصي باستخدام معامل ارتباط بايسيريال اذ ان رتبة تمييز الفقرة في هذا الاحصائي يكون اكثر استقرارا من عينة لأخرى عندما تتغير القدرة. ويمكن القول بأسلوب اخر ان قيمة معامل ارتباط بايسيريال المنخفضة لعينة من اي مستوى قدرة يشير إلى ان الفقرة منخفضة القدرة (أو عالية القدرة) قد تكون ببساطة دالة لصعوبة الفقرة وليست بالضرورة مؤشرا لضعف تمييزها.

4. وان كان مطور الاختبار يثق ودون تحيز بان العينات المستقبلية مشابهة في قدرتها لعينة التحليل، وكان الهدف اختيار فقرات ذات اتساق داخلي عالي فيفضل استخدام معامل ارتباط بوينت بايسيريال (Lord & Novik,1968). ومع ان هذا لم يتم اثباته بشكل نهائي الا انه يبدو منطقيا من خلال حقيقة ان قيم معاملات بوينت بايسيريال تكون اكبر للفقرات متوسطة الصعوبة، ومثل هذه الفقرات تسمح بأقصى تباين مشترك بين الفقرات وبالتالي اعلى قيمة لمعامل الفا.

5. وفي حالات التدرج الثنائي لكل من الفقرة والمحك، ينصح باستخدام معامل فاي أو تتراشورك، وحساب معامل فاي اسهل ولكنه محرر عندما تكون النسب غير متساوية لكلا المتغيرين. ويعتمد معامل تتراشورك على افتراضان درجات الفقرة والمحك تظهر عند تحويل متغيرين توزيعها اعتدالي إلى متغيرين ثنائيين. ونادرا ما ينصح باستخدام هذا المعامل كمقياس لتمييز

الفقرة للتعقيدات في حسابه، ولكن يوصي به بقوة عندما يراد استخدام الارتباطات الداخلية في التحليل العاملي.

حجم العينة:

لا توجد قاعدة مطلقة لأقل عدد من المفحوصين يلزم في تحليل الفقرات. ومن المؤكد ان تحليل الاختبارات مثل اختبارات الاستعداد أو التحصيل يستخدم بشكل واسع فيجب ان يعتمد على عينة ممثلة وكبيرة، وبالطبع الاف من المفحوصين. بالمقابل طالب الدكتوراه الذي يطور اداة لأطروحته يجب ان يعتمد على عينة اصغر من هذه بكثير وكقاعدة عامة يمكن حساب معظم المعالم الموصوفة في هذا الفصل بعينات مستقرة نسبيا ومؤلفة من (200) مفحوص، ويعد هذا اقل عدد مرغوب. وقانون اخر اثبت علميا هو ان يكون عدد المفحوصين من (5إلى10) اضعاف عدد الافراد (Nunnaly , 1967). وان استخدام مطور الاختبار هذا المؤشر فانه سيستخدم على الاقل (100) مفحوص لكل فقرة. وحجوم العينات اللازم لحساب معالم الفقرة في نظرية IRT يتراوح من (200 إلى 1000) مفحوص بالاعتماد على النموذج المختار في التحليل. ومن الضروري لمطور الاختبار ان يعرف طرائق الاختبار التي سيطبقها عندما يخطط للتطبيق الميداني، لذا فانه سيتم التطبيق على عدد كاف من المفحوصين.

2.2 الدراسات السابقة:

تبحث هذه الدراسة في المقارنة بين الطريقة الحدية والطرق الارتباطية لحساب تمييز الفقرة في ضوء تباين طول الاختبار وحجم العينة . الامر الذي تطلب الرجوع إلى جميع الدراسات – ما امكن – المتصلة بهما، ومن خلال اطلاع الباحثة على الدراسات السابقة بهذا الموضوع، تبين في حدود اطلاع الباحثة عدم وجود دراسات ذات صلة بموضوع الدراسة بشكل مباشر بل وجود دراسات تناولت اثر حجم العينة على الاختبارات وخاصة معلمة الصعوبة .

فمن الدراسات المبكرة في هذا السياق دراسة سيبرس (Cypress, 1973) هدفت لفحص أثر التواء العلامات وحجم العينة على تقديرات معلمة الصعوبة وقدرة الأفراد،

ولتحقيق الهدف من الدراسة استخدمت الباحثة خمس أحجام من العينات (600، 1200 ، 300 ، 150، 75) و 35 توزيعـا للعلامـات ميزهـا بسـبع مسـتويات للالتـواء وهـي: موجب عال، موجب متوسط، موجب متدن، سالب عالي، سالب متوسط، سالب متدن ومجموعة محكية توزيعها طبيعي، وتم إخضاع الجميع لاختبار في الرياضيات مكون من (90) فقرة. أشارت نتائج الدراسة بأنَّ الفروق بين معلمة القدرة للأفراد ذات التوزيع الملتوي ومعلمة القدرة للمجموعة المحكية (توزيع طبيعي) تزيد كلما زادت درجة الالتواء باتجـاه السـالب أو الموجـب، وبينـت النتـائج كـذلك بـأنَّ الأخطـاء المعياريـة لمعلمـة الصـعوبة كانـت متدنيـة عنـد التوزيعـات المتوسـطة وعاليـة عنـد التوزيعـات المتطرفـة بالالتواء.

وكشفت دراسة سونج(Seong, 1990) التي هدفت للتحقق من مدى حساسية تقـديرات معـالم الفقـرة لتوزيـع القـدرة باسـتخدام النمـوذج اللوجسـتي ثنـائي المعلمـة بـأنَّ تقديرات معلمة التمييز تكون أكثر دقة إذا كان شكل توزيع القدرة الحقيقة ملتويا التواءً سالباً وتوزيع شكل القدرة المسبق طبيعيا عندما يكون حجم العينـة صـغيراً، فـي حـين كانت تقديرات معلمة الصعوبة أكثر دقة عندما كان شكل توزيع القدرة الحقيقة ملتوياً التواءً موجباً وحجم العينة صغيراً.

وأجرت بـارنس ووايز (Barnes and wise , 1991) دراسـة بعنوان اسـتخدام النمـوذج أحـادي المعلمـة مـن نظريـة الاسـتجابة للفقـرة مـع العينـات الصـغيرة . فقـد اسـتخدمت عينـات صـغيرة الحجـم (50، 100، 200) فـرد وقـد تـم تطبيـق اختبـارين: الأول مكون من (25) فقرة و الثاني مكون من (50) فقرة، واستخدم نموذج راش) (Rasch Model فـي تقـدير معلمـة صـعوبة الفقـرة وقـدرات المفحوصـين، وتوصـلت الدراسـة إلـى أن تقدير قدرات المفحوصين يتـأثر بعدد الفقرات ولا يتـأثر بحجم العينـة، وأن تغير معلمة الصعوبة كان تغيراً بسيطاً عند تغير حجم العينـة، وقد يعود ذلك إلـى أن الفروق بين حجم العينات كان صغيراً.

وأجرى ستون (Stone, 1992) دراسـة هـدفت لتقيـيم التقـديرات الناتجـة عـن اسـتخدام طريقـة الارجحيـة العظمـى الهامشـية تحـت شـروط حجـم (Maximum Likelihood Estimaation) العينـة وطـول الاختبـار وتوزيـع طبيعـي ومفلطـح للقـدرة

باستخدام النموذج اللوجستي ثنائي المعلمة. كشفت نتائج الدراسة أنَّ تقديرات معلمة الصعوبة بقيت مستقرة عبر توزيعات القدرة ولكن تغيرت تقديرات معلمة التمييز عندما كان التوزيع المسبق للقدرة غير طبيعي وبقيت مستقرة عند التوزيع الطبيعي للقدرة.

وأجرى فان (Fan, 1998) دراسة هدفت لمقارنة النظرية التقليدية والنماذج ثنائية التدريج من حيث معالم الفقرات وقدرات الأفراد ومدى استقرار معالم الفقرات عبر العينات المختلفة. أشارت نتائج الدراسة إلى أنَّ درجة استقرار تقديرات النظرية التقليدية ونظرية الاستجابة للفقرة كانت الأكثر عبر العينات العشوائية واقل استقراراً عبر عينات القدرات المرتفعة والمتدنية، وبينت النتائج كذلك مع أن النموذج أحادي المعلمة اقل النماذج مطابقة للبيانات إلا أن تقديرات هذا النموذج كانت أكثر استقراراً من تقديرات النموذجين الثنائي والثلاثي المعلمة.

كما أجرى سوامنثان وزملاؤه (Swaminathan et al,2003) دراسة تحديد أثر الحكم المسبق على دقة تقديرات معلمات الفقرة في الاختبارات ثنائية التدريج باستخدام نظرية الاستجابة للفقرة، وقد هدفت الدراسة إلى تفحص أثر الطريقة الباييزية في تقدير معلمات الفقرة، واستقصاء العلاقة بين حجم العينة (Bayesian Estimation) ومواصفات المعلومات القبلية على دقة تقدير معلمات الفقرة وقد توصلت الدراسة إلى أن التحسن في تقدير معلمات الفقرة يختلف باختلاف النموذج اللوجستي (نموذج الترجيح اللوغاريتمي)؛ حيث تبين بأن النموذج اللوجستي أحادي المعلمة كان الأكثر تحسناً عند زيادة حجم العينة مقارنة بالنموذجين الثنائي والثلاثي .

واهتمت دراسة ساس وشمت ووكر (Sass, Schmitt & Walker, 2004)بدراسة أثر شكل التوزيع الملتوي للقدرة وحجم العينة وطريقة التقدير على معالم الفقرة ومعلمة القدرة للفرد وفق النموذج اللوجستي ثلاثي المعلمة، ولتحقيق الهدف من الدراسة استخدم الباحثون شكلين من توزيع القدرة ملتو(موجب، سالب) طبيعي وحجم عينة (500، 1000) وست طرق تقدير بينت النتائج بشكل عام بأنَّ التقديرات في معالم الفقرات كانت أقل دقة عندما كان توزيع القدرة ملتوي وأكثر دقة عندما كان توزيع القدرة طبيعي، وعند مقارنة الأخطاء المعيارية بين معلمة الصعوبة والتمييز لوحظ بأنها الأعلى عند معلمة التمييز.

وفي دراسة الثوابية (2010) التي هدفت إلى استقصاء أثر حجم العينة في تقدير معلمة صعوبة الفقرة والخطأ المعياري في تقديرها باستخدام نظرية الاستجابة للفقرة. ولتحقيق الهدف من الدراسة تم تقدير معلمة الصعوبة، والخطأ المعياري في تقديرها باستخدام اختبار في الرياضيات للصف العاشر الأساسي مكون من (80) فقرة وطبق على عينات عشوائية تراوح حجمها ما بين (200) طالب وطالبة إلى (11292) طالبا وطالبة.أشارت نتائج الدراسة بأن قيمة معلمة الصعوبة تزداد بزيادة حجم العينة، حيث بلغ متوسط صعوبة فقرات الاختبار 0.31 لوجيت عندما كان حجم العينة (200) وازداد بحيث أصبح (1.1)عندما كان حجم العينة (11292) طالبا وطالبة .وأشارت النتائج كذلك بان الخطأ المعياري في تقدير معلمة الصعوبة يتناقص بزيادة حجم العينة، حيث بلغ متوسط الأخطاء المعيارية في التقدير (0.32) عندما كان حجم العينة (200)، وتناقص بحيث أصبح (0.07) عندما كان حجم العينة (11292) طالبا وطالبة.

ومن هذه الدراسات دراسة الزبون (2013) التي هدفت إلى معرفة أثر حجم العينة على تقدير دالة معلومات الاختبار والخطأ المعياري في تقديرها باستخدام نظرية استجابة الفقرة، وللإجابة عن أسئلة الدراسة تم استخدام نتائج الاختبار الوطني لمادة الرياضيات، حيث تم اختيار عشوائيا اجابات(7500) طالبا وطالبة من طلبة الصف الثامن الأساسي على اختبار مكون من (40) فقرة من نوع الاختيار من متعدد، تم توزيعهم على خمس مجموعات بأحجام (500، 1000، 1500، 2000، 2500) مفحوص، وتم الاعتماد على برمجية (BILOG- MG3) وبرمجية(SPSS) لتحليل نتائج الاختبار الوطني لمادة الرياضيات، حيث تم إيجاد معالم الفقرات بالاعتماد على النموذج اللوجستي ثلاثي المعالم في نظرية استجابة الفقرة، وقد أظهرت النتائج أن مقدار دالة معلومات الاختبار تزداد بزيادة حجم العينة، كما أن الخطأ المعياري تقدير دالة معلومات الاختبار يتناقص بزيادة حجم العينة.

3.2 التعقيب على الدراسات السابقة:

في ضوء ما تقدم من دراسات حول الطريقة الحدية والطرق الارتباطية في حساب تمييز الفقرة ، تبين في حدود اطلاع الباحثة انه لا يوجد دراسات ذات صلة مباشرة لحساب التمييز في ضوء تباين طول الاختبار وحجم العينة، وأن معظم الدراسات تناولت اثر اختلاف حجم العينة وطول الاختبار على خصائص سيكومترية مثل صدق وثبات الاختبار وكذلك على معلمة الصعوبة والتخمين وذلك باستخدام النظرية الحديثة في القياس دون ان يكون هناك دراسة ذات صلة بمقارنة الطريقة الحدية والطرق الارتباطية في حساب تمييز الفقرة في ضوء تباين طول الاختبار وحجم العينة حسب النظرية الكلاسيكية في القياس، لذلك جاءت هذه الدراسة لتحقيق هذا الهدف.

الفصل الثالث

المنهجية والتصميم

يتناول هذا الفصل عرضاً لمنهجية الدراسة ومجتمعها وعينتها، والاداة التي تم استخدامها لجمع البيانات من اجل تحقيق اهداف الدراسة وخطوات التحقق من صدقها وثباتها واجراءات الدراسة وتصميمها والمعالجات الاحصائية التي استخدمت في تحليل واستخراج النتائج، والإجابة عن اسئلة الدراسة.

1.3 منهجية الدراسة:

استخدمت هذه الدراسة المنهج الوصفي التحليلي لتحقيق اهداف الدراسة والإجابة عن اسئلة الدراسة لاختبار تحصيلي في مبحث الرياضيات للصف التاسع الاساسي.

2.3 مجتمع الدراسة

تكون مجتمع الدراسة من طلاب الصف التاسع في مديرية تربية قصبة الكرك، وهم من طلاب المدارس الحكومية، وكان عددهم (1190) طالباً وطالبة ضمن اربع عشر مجموعة فرعية تتكون من احجام عينات مختلفة من خلال كشوفات مديرية التربية والتعليم في قصبة الكرك اذ تم الاعتماد على بيانات تم جمعها من قبل الطراونة(2017).

3.3 عينة الدراسة

تم اختيار عينة الدراسة باستخدام اسلوب المعاينة العشوائية البسيطة باستخدام برمجية spss ليصبح لدينا ثلاثة مجموعات من احجام العينة(150، 100، 50) طالباً، وكذلك ثلاث اطوال للاختبار (35، 25، 15) فقرة اذ تم الاعتماد على بيانات تم جمعها من قبل الطراونة(2017).

4.3 أداة الدراسة:

لتحقيق هدف الدراسة المقارنة بين الطريقة الحدية والطرق الارتباطية لحساب التمييز في ضوء تباين طول الاختبار وحجم العينة. تم استخدام اختبار تحصيلي للصف التاسع في مبحث الرياضيات والمكون بصورته النهائية من (35) فقرة محكم من قبل (10) من المحكمين المختصين، وقيمة معامل الثبات له (0.89)، وكذلك تم حساب ثبات الاعادة وكانت قيمته (0.84)، كما اعتمدت الباحثة (الطراونة، 2017) في كتابة الفقرات الأسس العامة المتبعة في بناء الاختبارات التحصيلية من قبل وزارة التربية والتعليم في مبحث الرياضيات والملحق(1) يبين الاختبار بصورته النهائية وفقاً للخطوات التالية.

اولاً: الاجراءات المستخدمة للتأكد من صدق الاختبار

صدق الاختبار:

تم التحقق من صدق المحتوى من خلال عرض الاختبار على مجموعة من المحكمين بلغ عددهم(10) محكماً منهم محكمين من تخصص القياس والتقويم في جامعة مؤتة، و(15) مدرساً ومشرفاً تربوياً في مديرية التربية واتعليم في محافظة الكرك، وطلب منهم إبداء وجهة نظرهم من حيث درجة تمثيل الاختبار للمحتوى، ودقة وسلامة الصياغة اللغوية، ووضوح الفقرات، وفي ضوء ملاحظاتهم تم التعديل على فقرات الاختبار بإعادة صياغة وحذف بعض الفقرات، حيث تكون الاختبار بصورته النهائية من (35) فقرة.

ثانياً: معامل التمييز والصعوبة:

تم حساب معامل تمييز الفقرات بحساب معامل الارتباط المصحح، حيث تم حساب معامل الارتباط بين درجة الطالب على الفقرة ودرجته الكلية على الاختبار مطروحاً منها درجته على الفقرة، وتم حساب معامل صعوبة كل فقرة من فقرات الاختبار وذلك بحساب المتوسط الحسابي لأداء طلبة العينة الاستطلاعية وعددهم(30) طالب على الاختبار والجدول (1) يبين النتائج الاستطلاعية كما ورد من الطراونة(2017).

جدول(1)

معاملات تمييز وصعوبة فقرات الاختبار

رقم الفقرة	معامل التمييز	معامل الصعوبة	رقم الفقرة	معامل التمييز	معامل الصعوبة	رقم الفقرة	معامل التمييز	معامل الصعوبة
1	0.41	0.70	13	0.34	0.63	25	0.65	0.67
2	0.39	0.47	14	0.53	0.60	26	0.58	0.27
3	0.46	0.40	15	0.35	0.73	27	0.63	0.40
4	0.38	0.63	16	0.62	0.40	28	0.58	0.27
5	0.31	0.57	17	0.37	0.27	29	0.57	0.30
6	0.34	0.60	18	0.37	0.22	30	0.52	0.30
7	0.42	053	19	0.34	0.60	31	0.56	0.23
8	0.31	0.60	20	0.43	0.33	32	0.40	0.30
9	0.42	0.80	21	0.34	0.34	33	0.62	0.40
10	0.65	0.47	22	0.42	0.40	34	0.55	0.23
11	0.34	0.60	23	0.54	0.33	35	0.38	0.33
12	0.57	0.7	24	0.60	0.30			

من خلال الجدول(1) يتضح بأن جميع قيم معاملات الصعوبة والتمييز لفقرات الاختبار كانت مقبولة، وقد تراوحت معاملات الصعوبة بين (0.22–0.80)، وتراوحت معاملات التمييز بين (0.31–0.65).

ثالثاً: ثبات الاختبار:

تم التحقق من ثبات الاختبار بطريقتين هما باستخدام ثبات الإعادة(Test Retest) حيث تم تطبيق الاختبار على عينة الدراسة الاستطلاعية والبالغة (22) طالب ورصدت درجاتهم، ثم أعيد تطبيق الاختبار على نفس افراد العينة مرة أخرى بعد (15) يوماً من التطبيق الأول ورصدت درجاتهم، وبحساب معامل الارتباط بين درجات الطلبة بين مرتي التطبيق بلغ معامل الارتباط المحسوب بهذه الطريقة(0.84)، كما تم استخدام معادلة كيودر ريتشاردسون(20) للتحقق من دلالات صدق الاتساق الداخلي وبلغ معامل الثبات المحسوب بهذه الطريقة(0.89)، وتعد هذه القيمة جيدة لمثل هذا النوع من الاختبارات، مما يشير إلى أن الاختبار يتمتع بدرجة ثبات مناسبة.

5.3 اجراءات الدراسة:

تم إجراءات الدراسة ضمن خطوات من أجل الإجابة عن اسئلة الدراسة وبالصورة الاتية:

أولاً: تم اختيار حجم العينة وتقسيمها إلى ثلاث مجموعات بأحجام مختلفة وهي (150، 100، 50) مفحوص، وكذلك ثلاثة أطوال للاختبار وهي (35، 25، 15) فقرة، وجميعها تم اختيارها بالطريقة العشوائية باستخدام برمجية spss .

ثانياً: تم حساب معامل التمييز للفقرات بالطرق الثلاث المستخدمة بالدراسة (الطريقة الحدية، معامل ارتباط بوينت بايسريال، معامل ارتباط فاي) حيث تم حساب التمييز بالطريقة الحدية ومعامل ارتباط بوينت بايسريال باستخدام برمجية الاكسل (2010) للأحجام المختلفة لعينات الدراسة المستخدمة، وكذلك وفقاً للأطوال المختلفة للاختبار وطريقة فاي باستخدام (spss).

ثالثاً: تم حساب المتوسط الحسابي للتمييز حسب طريقة من الطرق الثلاث (الطريقة الحدية، معامل ارتباط بوينت بايسريال، معامل ارتباط فاي) وفقاً لحجم العينة وطول الاختبار والمبينة في الجدول(2).

رابعاً: تم استخدام تحليل التباين الثنائي لمعرفة أثر طول الاختبار وحجم العينة وأثر التفاعل بينهم على متوسطات معامل التمييز والموضحة في الجدول (4) باستخدام برمجية spss.

الفصل الرابع

عرض النتائج ومناقشتها والتوصيات

يتضمن هذا الفصل عرضا للنتائج التي تم التوصل إليها من خلال هذه الدراسة، والتي هدفت إلى المقارنة بين الطريقة الحدية والطرق الارتباطية في حساب تمييز الفقرة في ضوء تباين طول الاختبار وحجم العينة باستخدام (3)اختبارات كل اختبار مكون من عدد من الفقرات الاختبار الأول مكون من (15) فقرة والثاني مكون من (25) فقرة والثالث مكون من(35) فقرة وبأحجام عينات مقسمة إلى (50، 100، 150) طالب في مبحث الرياضيات للصف التاسع ؛ بالإضافة إلى مناقشة هذه النتائج وتفسيرها، وقد تم عرض هذه النتائج بحسب اسئلة الدراسة ومتغيراتها حسب منهجية منظمة تقوم على عرض السؤال ثم نوع الإحصائي المناسب، يلي ذلك جدول البيانات ثم التعليق عليها.

1.4 عرض النتائج ومناقشتها

النتائج المتعلقة بالسؤال الأول والثاني الذي ينص على: " هل تختلف قيم مؤشرات تمييز الفقرة حسب طرق حساب مؤشر التمييز تبعاً لطول الاختبار وحجم العينة؟

للإجابة عن السؤال الأول، تم حساب قيمة معامل التمييز بثلاثة طرق (الطريقة الحدية، معامل ارتباط بوينت بايسيريال، معامل ارتباط فاي) الطريقة الأولى الحدية تم حسابها باستخدام برمجية الاكسل وكذلك معامل ارتباط بوينت بايسيريال اما معامل ارتباط فاي فتم حسابة عن طريقة برمجية spss ومن ثم تم حساب متوسطات معامل التمييز للثلاثة طرق و الجدول (2) يبين ذلك.

جدول (2)

الوسط الحسابي لمعامل التمييز باختلاف طول الاختبار وحجم العينة

حجم العينة	طول الاختبار	التمييز الحدية	معامل ارتباط بوينت بايسيريال	معامل فاي
50	15	0.25	0.35	0.57
	25	0.27	0.40	0.52
	35	0.27	0.38	0.49
100	15	0.26	0.32	0.71
	25	0.23	0.30	0.49
	35	0.23	0.30	0.43
150	15	0.21	0.29	0.62
	25	0.23	0.29	0.51
	35	0.22	0.28	0.44

يتضح من النتائج الواردة في الجدول(2) تباينا ملحوظاً في الأوساط الحسابية لقيم مؤشر التمييز تبعا لطول الاختبار وحجم العينة، حيث كان أعلى وسط حسابي له باستخدام معامل فاي عندما كان حجم العينة يساوي (50) وطول الاختبار(15) كانت قيمة معامل التمييز(0.57)، وعند حجم عينة (100) كانت قيمة معامل التمييز(0.71) وطول اختبار(15)، وكذلك عند حجم عينة (150) وطول اختبار(15)، كان معامل التمييز(0.62). اما بالنسبة لمعامل الارتباط بوينت بايسيريال كان اعلى وسط حسابي له عند حجم عينة(50) وطول الاختبار(25)، كان معامل التمييز(0.40)، كذلك عند حجم عينة(100) وطول اختبار(15)، كان معامل التمييز(0.32)، وكذلك نجد ان قيمة معامل الارتباط بوينت بايسيريال عند حجم العينة(150)، وطول الاختبار(15،25)، كانت قيمة معامل التمييز(0.29). واخيراً الطريقة الحدية كان اعلى متوسط حسابي لها عند حجم عينة(50) وطول اختبار(25،35)،بينما عند حجم عينة (100) كان اعلى قيمة لمتوسط معامل التمييز هي (015) عند طول اختبار(15)، وكذلك نجد ان اعلى متوسط لقيمة معامل التمييز الحدي عند حجم عينة (150) كان قيمته (0.23) عند طول اختبار(25) فقرة.

وللإجابة عن سؤال الدراسة الثاني والثالث حول تأثير حجم العينة وطول الاختبار

على معامل التمييز واثر التفاعل بينهما اجرينا تحليل التباين الثنائي والتي كانت نتائجه كما يلي:

أولاً التحليل الاحصائي الوصفي للمتوسطات الحسابة والانحراف المعياري كما يظهر في الجدول(3)

الجدول (3) التحليل الوصفي للبيانات

حجم العينة	طول الاختبار	الوسط الحسابي	الانحراف المعياري
50	15	.3927	.16189
	25	.4285	.24611
	35	.3742	.21797
	المجموع	.3985	.18478
100	15	.3996	.12219
	25	.3424	.13420
	35	.3450	.15323
	المجموع	.3624	.12202
150	15	.3801	.10562
	25	.3216	.10878
	35	.3142	.11064
	المجموع	.3387	.09892
المجموع	15	.3908	.11466
	25	.3642	.15815
	35	.3445	.14657
	المجموع	.3665	.13684

يتضح من الجدول (3) ان اعلى متوسط حسابي لمعاملات التمييز عند حجم العينة (50) وطول الاختبار (25) حيث بلغت قيمته(0.4285)، بينما اقل قيمة كانت(0.3142) كانت عند طول الاختبار(35) وحجم عينة(150). وعند استعراض المتوسطات الحسابية لمعامل التمييز باختلاف حجم العينة وطول الاختبار نجد ان هناك تفاوت بسيط بينها. على العكس نجد ان الانحراف المعياري للأوساط الحسابية لمعاملات التمييز يتناقص بزيادة حجم العينة وطول الاختبار ولمعرفة التأثير الحقيقي لكل من حجم العينة وطول الاختبار واثر التفاعل بينهما والجدول (4) يبين ذلك.

الجدول (4)

نتائج التحليل التباين الثنائي للكشف عن دلالة الفروق بين متوسطات مؤشر التمييز حسب متغيري طول الاختبار وحجم العينة

المعلمة	مصدر التباين	مجموع المربعات	درجات الحرية	وسط المربعات	قيمة F	الدلالة الاحصائية
التمييز	حجم العينة	.016	2	.008	.325	.727
	طول الاختبار	.010	2	.005	.194	.825
	التفاعل	.009	4	.002	.089	.985
	الخطأ	.452	18	.025		
	الكلي	4.114	27			

يتضح من الجدول (4) ان حجم العينة ليس له تأثير معنوي على معامل التمييز حيث كانت قيمة الدلالة الاحصائية (0.727) وهي اكبر من القيمة (5%) وبناء عليه نقبل الفرض الصفري الذي ينص على "أنه لا يوجد تأثير لحجم العينة على مؤشر التمييز حسب النظرية التقليدية في القياس"، وكذلك نجد ان طول الاختبار بلغت قيمة الدلالة الاحصائية (0.825) وهي اكبر من (5%) وهذا يدل على انه لا يوجد تأثير ايضا لطول الاختبار على معلمة التمييز وانها لا تختلف باختلاف طول الاختبار من ناحية احصائية، وأما بالنسبة لأثر التفاعل بين حجم العينة وطول الاختبار فنجد ان قيمه الدلالة الاحصائية (0.985) وهي تدل على عدم وجود اثر لأثر التفاعل بينهما. ومن ذلك نستنتج انه لا يوجد تأثير لحجم العينة وطول الاختبار والتفاعل بينهما على معلمة التمييز ضمن هذه الاطوال وحجم العينات في هذه الدراسة وفق معادلات النظرية الكلاسيكية في القياس ويكمن تفسير ذلك بعدم وجود فروق ذات دلالة احصائية بين الطرق المستخدمة في الدراسة لأنها اتفقت مع الأدب السيكومتري الذي اعطى بدائل لحساب تمييز الفقرة وهذه البدائل تعطي نفس النتيجة وبالتإلى يمكن استخدام أي من هذه الطرق في حساب معامل التمييز.

2.4 التوصيات

1- إجراء دراسات على الفقرات ذات التصحيح الجزئي الاسئلة الانشائية للمقارنة بين معامل التمييز.

2- استخدام اي من هذه الطرق يؤدي الى نفس النتيجة في الحكم على الفقرة .

3- اجراء دراسات اخرى لحساب معامل التمييز باستخدام احجام عينات وطول اختبار مختلفة.

قائمة المراجع

أ. المراجع العربية

أبوحطب، فؤاد (1992) . **دليل المعلم في تقويم الطالب** . المركز القومي للامتحانات والتقويم التربوي بالاشتراك مع وزارة التربية والتعليم، القاهرة: دار غريب للطباعة.

الثوابية، أحمد محمود (2010).**أثر حجم العينة على تقدير صعوبة الفقرة والخطأ المعياري في تقديرها باستخدام نظرية الاستجابة للفقرة**.مجلة جامعة دمشق، 26 -، 225، 556.

جردات، عزت، عبيدات، ذوقان، وابو غزالة، هيفاء (2002). **مبادئ القياس والتقويم ط3**. الاردن، عمان: جهينة للنشر.

الحجاج، ايمن(2013).اثر ترتيب الفقرات حسب خصائصها السيكومترية على أداء الافراد والخصائص السيكومترية للاختبار. الاردن، جامعة مؤتة. رسالة ماجستير غير منشورة.

خوري، توما جورج (2008). **القياس والتقويم في التربية والتعليم**. لبنان، بيروت: مجد المؤسسة الجامعية للدراسات والنشر والتوزيع.

الزبون، حابس (2013)، **أثر حجم العينة على تقدير دالة المعلومات للاختبار والخطأ المعياري في تقديرها باستخدام النظرية الحديثة في القياس**، مجلة جامعة النجاح للأبحاث (العلوم الإنسانية)، المجلد 27(6).

الطراونة، ناديا(2017). **أثر حجم العينة في تقدير معامل الثبات في ضوء نظرية التعميم**. الاردن. جامعة مؤتة. رسالة ماجستير غير منشورة.

الظاهر، واخرون(2002). **مبادئ القياس والتقويم في التربية. الاردن**، عمان: الدار العلمية الدولية للنشر والتوزيع ودار الثقافة للنشر والتوزيع.

عبدالحافظ، شحته عبدالمولى (1999).**تقويم بناء الاختبارات المرجعة إلى محك / المعيار في نظرية الاستجابة للمفردة والنظرية التقليدية.** رسالة دكتوراه غير منشورة ، كلية التربية جامعة عين شمس : مصر.

عبدالهادي، نبيل (2002). **مدخل إلى القياس والتقويم واستخدامه في مجال التدريس الصفي ط2.** الاردن، عمان: دار النشر.

علام، صلاح الدين محمود (2001).**الاختبارات التشخيصية مرجعية المحك في المجالات التربوية والنفسية والتدريبية.** ط 2، القاهرة: دار الفكر العربي

علام، صلاح الدين محمود (2005).**نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي.** القاهرة: دار الفكر العربي.

علام، صلاح الدين محمود(2007). **القياس والتقويم التربوي في العملية التدريسية.** الاردن، عمان: دار المسيرة للنشر والتوزيع.

عودة، احمد (2010). **القياس والتقويم في العملية التدريسية ط3.** الاردن، اربد: دار الامل للنشر والتوزيع.

النجار، نبيل (2010). **القياس والتقويم: منظور تطبيقي مع تطبيقات برمجية** spss. **الاردن،** عمان: دار الحامد للنشر والتوزيع.

ب. المراجع الأجنبية

Barnes, L.and Wise (1991). **The utility of a modified one - parameter IRT model with small samples**, Journal of Applied Measurement in Education, (4)(2),143. – 153.

Beuchert .A.K. and Mendoza .J .L(1979). **A Monte Carlo comparison of ten item discrim-ination indices**. Journal of Educational Measurement.16.109-118.

Crocker, L. ; Algina, J. (1986**). Introduction to Classical and Modern Tests Theory**. New York: CBS College Publishing.

Cypress, k.(1973).**The effect of diverse test score distribution characteristics on the estimation of Rash measurement model**. Paper presented at the Annual Meeting of Americans Educational Research Association, Orleans.

Englehart,M.D.(1965).**A comparison of several item discrimination indices**. Journal of educational Measurement,2,69-76.

Fan, X.(1998). **Item response theory and classical test theory. An empirical comparison of their item /person statistics**. Educational and Psychological Measurement, 58, 257-381.

Findley, W.G (1956).**A rationale for evaluation of item discrimination statistics**. Educational and Psychological Measurement, 16,175-180.

Hambelton, R. ; Swaminathan, S. & Rogers, H. J. (1991). **Fundamentals of Item Response Theory**. SAGE publications, Newbury Park: The international professional publishers.

Hambelton,R.K.(1983). **Item response theory, Boston, Kluwer**. Nijhalf puplishing.

Kelley .T .L .(1939).**Selection of upper and lower groups for the validation of test items**. Journal of Educational Psychology,30.17-24.

Lane, D.S, and Bull, K.S.(1987). **The effects of knowledge of item arrangement, Genderm, and statistical and cognitive item difficulty on test performance**. Educational and Psychological.47

Lord. F.M. (1968).**An analysis of the verbal scholastic Aptitude Test using Birnbaums three- parameter logistic model**. Educational and psychological Measurement .28,989-1020.

Lord.F.M. and Novick .M.R.(1968).**Statistical theories of mental test scores.Reading**.Mass:Addison-Wesely.1968.

Nunnally ,J.C. ?(1967).**Psychometric theory .New York: McGraw-Hill** .Osgood ,C, Suci ,G ,and Tannenbaum .P.(1957).The measurement of meaning .Urbana. University of 111inois press

Oosterhof ,A.C.(1976).**Similarity of various item discrimination indices** .Journal of Educational Measurement,13,145-150.

Plake, B.S, Thompson, P.A, and lowryms.(1981). **Efects of item arrangement, and test anxiety on tow scoring methods**. The journal experimental education.44.

Sass, D. A., Schmitt, T. A., & Walker, C. M.(2004). **An evaluation of BILOG-MG with skewed theta distribution using various estimation procedures : A simulation study**. Poster presented at the National Council on Measurement in Education. San Diego, California.

Seong, T. J.(1990). **Sensitivity of Marginal Maximum Likelihood Estimation of item and ability parameter to the characteristics of the prior ability distribution**. Applied Psychological Measurement,14,299-311.

Stone,C. A.(1992). **Recovery of Marginal Maximum Likelihood Estimation in Two Parameter Logistic Response Model: An evaluation of MULTILOG.** Applied Psychological Measurement ,16, 1-16.

Swaminathan.H et al. (2003). **Small sample estimation in dichotomous item response models: Effect of priors based on judgmental information on the accuracy of item parameter estimates**, Journal of Applied Psychological Measurement, 27 (1), 27-51.

الملاحق

الملحق (أ)

أداة الدراسة بصورتها النـهائيـة

(1) طول قطر الدائرة التي معادلتها س^2 + ص^2 = 64 يساوي

أ) 8 ب) 16

جـ) 4 د) 64

(2) احـداثيا منتصـف القطعـة المسـتقيمة أب ، حيـث أ (− 6 ، − 8) ، ب (2 ، 4) يساوي:

أ) (− 2 ، − 2) ب) (− 4 ، − 6)

جـ) (− 2 ، 2) د) (3 ، 2)

(3) طول القطعة المستقيمة س ص حيث س (0 ، 0) ، ص (− 3 ، 4) يساوي :

أ) 25 ب) 16

جـ) 5 د) 7

(4) إحدى النقاط المذكورة أدناه تقع على الدائرة التي معادلتها س^2 + ص^2 = 25

أ) (− 4 ، 2) ب) (− 3 ، 4)

جـ) (4 ، 0) د) (5 ، 3)

(5) معادلة الدائرة التي مركزها (− 2 ، 5) وطول نصف قطرها 2 هي :

أ) (س + 2)^2 + (ص − 5)^2 = 4 ب) (س + 2)^2 + (ص − 5)^2 = 2

جـ) (س − 2)^2 + (ص − 5)^2 = 2   د) (س − 2)^2 + (ص − 5)^2 = 4

(6) ما احداثيا مركز الدائرة التي معادلتها س^2 + ص^2 − 6 س + 8 س − 10 = 0 :

أ) (3 ، − 4)   ب) (− 3 ، 4)

جـ) (6 ، − 8)   د) (− 6 ، 8)

(7) أي النقاط الآتية تقع على محيط الدائرة التي معادلتها س^2 + (ص^2 − 2) = 25

أ) (5 ، 0)   ب) (4 ، 5)

جـ) (0 ، 4)   د) (0 ، 4)

(8) معادلة الخط المستقيم الذي ميله (5) ويمر بنقطة الأصل هي :

أ) ص = 5   ب) ص = 5 س

جـ) ص = س + 5   د) س = 5 ص

(9) أي المعادلات الآتية تمثل معادلة دائرة

أ) س^2 + ص = 25   ب) س^2 − ص^2 = 25

جـ) س^2 + 4 ص^2 = 25   د) س^2 = 25 − ص^2

(10) ميل الخط المستقيم الذي معادلته (ص − 3) 2 (3 − س) يساوي :

أ) 3   ب) − 3

جـ) 2   د) − 2

(11) إذا كانت النقطتان هـ (1 ، 3) ، و (3 ، − 1) نقطتين في المستوى الاحداثي ، وكانت النقطة هـ نقطة منتصف القطعة المستقيمة و ل، فما احاثيا النقطة ل

أ) (2 ، 1)   ب) (− 1 ، 7)

جـ) (4 ، 2)   د) (5 ، − 5)

(12) إذا كانت س قياس زاوية حادة ، وكان جا س = 0,6، وكانت س زأوية حادة، فان ظا س يساوي :

أ) 0,75   ب) 0,6

جـ) 0,4   د) 0,8

(13) إذا كان جا س = 3 جتا س، وكانت س زاوية حادة، فان ظا س يساوي :

أ) 3 ب) $\frac{1}{3}$

جـ) 2 د) $\frac{1}{2}$

(14) إذا كان س قياس زاوية حادة ، فان إحدى العبارات التإليه صحيحة :

أ) جا س > 1 ب) جا س < 0

جـ) 0 < جا س < $\frac{1}{2}$ د) 0 < جا س < 1

(15) إذا كان جتا 5 س = 0,8546 حيث 0° < 5 س < 90° ، فان قيمة المقدار جا (90° − 5 س) − جتا 5 س =

أ) صفر ب) 0,1454

جـ) 1 د) 0,8546

(16) القيمة العددية للمقدار : $\frac{\text{جا}30}{\text{جتا}60} + \text{ظا}45$ **يساوي :**

أ) 5 ب) 3

جـ) 2 د) 1

(17) إذا كان ظا س = 5 ، فان ظا (90 − س) يساوي :

أ) 5 ب) 0,75

جـ) 1 د) $\frac{1}{5}$

(18) أ ب ج مثلث قائم الزاوية في ب فيه (أ ب = 3 سم) ، (ب ج = 4 سم) ، فان جا أ يساوي

أ) $\frac{4}{5}$ ب) $\frac{3}{5}$

جـ) $\frac{4}{3}$ د) $\frac{3}{4}$

(19) إذا كـان الوسـط الحسـابي لخمـس مشـاهدات هـي 10 ، 12 ، س ، 17 ، 15، يساوي 15 ،فان قيمة المشاهدة س تساوي :

أ) 15 **ب) 13**

جـ) 21 **د) 52**

(20) الفئة (10 – 20) تحتوي العدد :

أ) 50 **ب) 100**

جـ) 15 **د) 75**

(21) صف فيه 20 طالبا ، إذا كان مجموع علامات أول 7 طلاب فيه يساوي 70، ومجموع علامات بقية الطلبة يساوي 110 . فان الوسط الحسابي لعلامات الصف هو :

أ) 9 **ب) 18**

جـ) 20 **د) 13,5**

(22) إذا كانت القيم الآتية مرتبة تصاعديا 7 ، 9 ، س ، 15 ، 23 ، 30 ، وكان الوسيط لتلك القيم يساوي 14 ، فان قيمة س هي :

أ) 15 **ب) 13**

جـ) 14 **د) 14,5**

(23) وعاء يحتوي على 5 كرات حمراء و3 كرات زرقاء وكرتين خضراء ، إذا تم سحب كرة عشوائيا من الوعاء ، فان حادثا من الحوادث الآتية يعد مستحيلا :

أ) اختيار كرة حمراء **ب) اختيار كرة زرقاء**

جـ) اختيار كرة صفراء **د) لا شيء مما ذكر**

(24) احتمال اختيار عدد أولي من مجموعة الأعداد : { 19 ، 17 ، 13 ، 11 ، 7 ، 5 ، 3 ، 2 }

أ) 1 **ب) صفر**

جـ) $\frac{1}{8}$ **د)** $\frac{7}{8}$

(25) واحدة من التجارب الآتية تمثل تجربة عشوائية :

أ) تجربـة رمـي قطعة نقد

ب) اختيـار كـرة من وعاء به كرات مختلفة اللون دون النظر إلـيها

جـ) تجربة رمي حجر نرد

د) جميع ما ذكر

(26) عدد الطرق التي يستطيع بها شخص تكوين كلمة مكونة من الحروف الثلاثـة { ح ، ل ، أ } دون أن يكرر الحرف أكثر من مرة في الكلمة الواحدة :

أ) 3 طرق

ب) طريقة واحدة

جـ) 6 طرق

د) 9 طرق

(27) ألقيت قطعة نقد معدنية 50 مرة، فظهرت الكتابـة 20 مرة، الاحتمـال التجريبي لظهور الصورة هو :

أ) 0,6

ب) 0,4

جـ) 0,5

د) 0,15

(28) القي حجر نرد مرتين متتإلىتين . احتمال الحصول على عددين مجموعهما 5 هو :

أ) $\frac{1}{6}$

ب) $\frac{5}{36}$

جـ) $\frac{1}{8}$

د) $\frac{1}{9}$

(29) قيمة س التي تحقق المعادلة $3^{س - 1} = 27$ تساوي :

أ) 3

ب) 4

جـ) 1

د) 2

(30) المقدار ($7 \times 10^{-1} + 3 \times 10 + 4 \times 10^{2}$) هو تحليل للعدد :

أ) 430,7

ب) 43,7

جـ) 437

د) 43,07

(31) إحدى العبارات الرياضية الآتية صحيحة :

أ) $س^3 \times س^2 = س^6$

ب) $س^3 + س^2 = س^6$

جـ) $س^3 \div س^2 = س^5$

د) $س^3 \times س^2 = س^5$

(32) مجموعة الحل للمعادلة الأسية $4^{ص} \times 8^{2ص} = 256$ هي :

أ) {– 8 } ب) { – 1 }

جـ) { 1 } د) { 8 }

(33) أي مما يأتي أفضل تقريب ل 12,9 × 3,09 :

أ) 12 × 3 ب) 13 × 3

جـ) 13 × 4 د) 12 × 4

(34) الزوج المرتب الذي يحقق كلا من المعادلتين الآتيتين معا 2 س + ص = 5 ، س – ص 4 هو :

أ) (3 ، – 1) ب) (– 3 ، – 1)

جـ) (– 3 ، 1) د) (3 ، 1)

(35) إذا كان (س + 1 ، – 6) = (2 ، 4 ص – 3) ، فان قيمة س ، ص تساوي :

أ) س = 3 ، ص = 3 ب) س = 1، ص = – 1

جـ) س = – 3 ، ص = 3 د) س = – 1 ، ص = 1

المعلومات الشخصية

الاسم: اخلاص جميل خلف الصرايرة

التخصص: قياس و تقويم

الكلية: العلوم التربوية – قسم علم النفس

سنة التخرج: 2017

الايميل: ekhlass.1975@yahoo.com